

# EXHIBIT D

**UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,  
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,  
OPENAI OPCO LLC, OPENAI GLOBAL LLC,  
OAI CORPORATION, LLC, and OPENAI  
HOLDINGS, LLC,

Defendants.

Civil Action No. 1:23-cv-11195-SHS

**THE NEW YORK TIMES COMPANY’S SECOND SET OF REQUESTS FOR  
PRODUCTION TO OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,  
OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, AND  
OPENAI HOLDINGS, LLC**

Pursuant to the Federal Rules of Civil Procedure and the Local Rules of the Southern District of New York, (the “Local Civil Rules”), The New York Times Company (“The Times”) requests that OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC (“OpenAI Defendants”) produce the requested documents at the offices of Susman Godfrey L.L.P., One Manhattan West, New York, New York 10001-8602, within thirty days after the date of service of this document, or at such other mutually agreeable time and place.

**I. DEFINITIONS**

The following definitions and instructions apply to the requests listed below:

1. All definitions and rules of construction set forth in Rule 26.3 of the Local Rules of the Southern District of New York are incorporated by reference herein.

8. If any documents requested herein have been lost, discarded or destroyed, they shall be identified as completely as possible, including, without limitation, the following information: (a) the date of disposal; (b) the manner of disposal; (c) the reason for disposal; (d) the person authorizing disposal; and (e) the person disposing of the document.

9. If no Documents responsive to a particular request exist, state that no responsive Documents exist.

### **III. DOCUMENT REQUESTS**

#### **REQUEST FOR PRODUCTION NO. 16:**

Documents concerning the decision not to release Training Datasets or other information for Defendants' Text Generation AI Model(s) after GPT-3 (including all versions thereof, such as GPT-4 Turbo and GPT-4o), including any communications regarding this decision and concerns regarding the effects of releasing this data on safety, ethics, profitability, competition, and legal obligations and/or ramifications.

#### **REQUEST FOR PRODUCTION NO. 17:**

Documents concerning the compilation, acquisition, and curation of Training Datasets, including: (i) communications regarding paying for or obtaining permission to use any Training Datasets or content therein, ii) communications with Training Dataset creators, including Common Crawl, and iii) communications concerning the relevance and impact of paywalls on building Training Datasets or gathering training data.

#### **REQUEST FOR PRODUCTION NO. 18:**

Documents concerning the removal of any content, including information regarding authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use of works, from Training Datasets before, during, or after training of Defendants' Generative AI Model(s).

Defendants' Generative AI Products and Services contains Times Content, (iii) any methods, procedures, or instructions to avoid overfitting or memorization of Times Content in Defendants' Generative AI Models, (iv) any differences in the foregoing among Defendants' Generative AI Products and Services, and (v) any changes in the foregoing over time.

**REQUEST FOR PRODUCTION NO. 29:**

Documents concerning Defendants' reliance on earlier versions of Defendants' Generative AI Models to be used in or to train or develop later versions of Defendants' Generative AI Models, including Defendants' use of outputs, weights, or any learnings acquired from earlier versions of Defendants' Generative AI Models.

**REQUEST FOR PRODUCTION NO. 30:**

The full model cards, data set cards, or system cards for each of Defendants' Generative AI Models, including those not released to the public.

**REQUEST FOR PRODUCTION NO. 31:**

Documents relating to Your technical documentation for Defendants' Generative AI Models, Products, and Services as defined in Annex IV of the European Union's Artificial Intelligence Act.

**REQUEST FOR PRODUCTION NO. 32:**

Documents concerning the training curricula, stages, iterations, cycles, batches, or epochs for Defendants' Generative AI Models.

**REQUEST FOR PRODUCTION NO. 33:**

Documents concerning Defendants' decisions regarding which works to use to train and develop Defendants' Generative AI Models, and how to use those works in training, including (i) documents concerning the factors considered in determining the weighting or frequency at which

works are seen during the training of Defendants' models, (ii) documents concerning de-duplication of Training Datasets; (iii) documents concerning any classifiers and/or filters applied to Training Datasets that determine document quality and/or remove low quality documents; (iv) documents concerning rules and procedures for determining sample frequency; (v) documents sufficient to show how the training weights or parameters within Defendants' Generative AI Models would change if Times Content was not used during training; and (vi) documents sufficient to show which individuals made or make decisions regarding which works to use to train and develop Defendants' Generative AI Models, and how to use those works in training.

**REQUEST FOR PRODUCTION NO. 34:**

Documents sufficient to describe each aspect or phase of the process for training Defendants' Generative AI Models, including the pre-training, training, fine-tuning, and any other refinement or development processes as well as any distinctions or differences between and among the aspects or phases.

**REQUEST FOR PRODUCTION NO. 35:**

Documents concerning any changes made to Defendants' Generative AI Models or Generative AI Products and Services from 2021 to present, including any logs showing changes, documents relating to ablation or "unlearning," (i.e. the ability to remove memorization, content, or other effects of specific parts of the Training Datasets from the models), and documents and communications concerning any changes to the models made or considered, including in response to claims of copyright infringement or improper use of data.

**REQUEST FOR PRODUCTION NO. 36:**

Documents concerning the differences between Defendants' Generative AI Models, including (i) differences in the Training Datasets and/or the composition of the Training Datasets

used for each model; (ii) differences in the training curricula used for each model; and (iii) the differences between the “Turbo” and non-Turbo versions of Defendants’ models.

**REQUEST FOR PRODUCTION NO. 37:**

The following documents concerning each of Defendants’ Generative AI Models, Products, and Services: (i) product briefs, (ii) readme files, (iii) product requirement documents, (iv) product or project management tools and records (e.g., Jira), (v) documentation from workflow management programs, (vi) minutes or notes from technical team meetings, (vii) documentation concerning data loaders, (viii) documents related to red-teaming and/or stress testing, (ix) technical system or architecture documents, specifications, or reports, and (x) launch documents.

**REQUEST FOR PRODUCTION NO. 38:**

Source code related to (i) Defendants’ Generative AI Models, Products, and Services, (ii) training curricula for Defendants’ Generative AI Models, (ii) any ranking, weighting, or scoring of training data, (iii) data loaders, (iv) Your web crawlers, spiders, and bots, including GPTBot, and (v) any processing of training data.

**REQUEST FOR PRODUCTION NO. 39:**

Documents concerning the study, testing, or evaluation of Defendants’ Generative AI Products and Services and Generative AI Models for hallucinations, memorization, and any problematic behaviors, including: (i) the results of the efforts of safety teams, including any red-teaming efforts, (ii) responses to ChatGPT’s feedback contest, (iii) submissions to Your bug bounty program, (iv) any other red-teaming, and (v) the extent to which rates of hallucination and memorization have changed over time within Defendants’ Generative AI Models.

**REQUEST FOR PRODUCTION NO. 40:**

Documents concerning the prevalence of outputs that are verbatim copies of text included in Training Datasets, including any documents concerning (i) steps taken to prevent or encourage

**REQUEST FOR PRODUCTION NO. 69:**

Documents concerning Your policies or practices regarding public access to Your governing documents, financial statements, and conflict of interest rules.<sup>6</sup>

**REQUEST FOR PRODUCTION NO. 70:**

Documents concerning referral, affiliate, advertising, and paid search revenue generated by Defendants' Generative AI Products and Services.

**REQUEST FOR PRODUCTION NO. 71:**

Documents sufficient to show in detail, on a monthly, quarterly, or annual basis, the revenue and profit realized from Defendants' Generative AI Products and Services, including documents sufficient to show in detail, by product or service, the gross revenue generated, the costs, the gross profit margin, and the net profit margin.

**REQUEST FOR PRODUCTION NO. 72:**

Documents relating to the pricing of Defendants' Generative AI Products and Services in the United States, including price quotes, price lists, subscription amounts, price announcements, and actual prices paid by customers.

Dated: May 22, 2024

/s/ Zach Savage

Ian Crosby (*pro hac vice*)  
Genevieve Vose Wallace (*pro hac vice*)  
Katherine M. Peaslee (*pro hac vice*)  
SUSMAN GODFREY L.L.P.  
401 Union Street, Suite 3000  
Seattle, WA 98101  
Telephone: (206) 516-3880  
Facsimile: (206) 516-3883  
icrosby@susmangodfrey.com

---

<sup>6</sup> Paresh Dave, *OpenAI Quietly Scrapped a Promise to Disclose Key Documents to the Public*, WIRED (Jan. 24, 2024), <https://www.wired.com/story/openai-scrapped-promise-disclose-key-documents/>.

gwallace@susmangodfrey.com  
kpeaslee@susmangodfrey.com

Davida Brook (*pro hac vice*)  
Emily K. Cronin (*pro hac vice*)  
Ellie Dupler (*pro hac vice*)  
SUSMAN GODFREY L.L.P.  
1900 Ave of the Stars, Suite 1400  
Los Angeles, CA 90067  
Telephone: (310) 789-3100  
Facsimile: (310) 789-3150  
dbrook@susmangodfrey.com  
ecronin@susmangodfrey.com  
edupler@susmangodfrey.com

Elisha Barron (5036850)  
Zachary B. Savage (ZS2668)  
Tamar Lusztig (5125174)  
Alexander Frawley (5564539)  
Eudokia Spanos (5021381)  
SUSMAN GODFREY L.L.P.  
One Manhattan West  
New York, NY 10001  
Telephone: (212) 336-8330  
Facsimile: (212) 336-8340  
ebarron@susmangodfrey.com  
zsavage@susmangodfrey.com  
tlusztig@susmangodfrey.com  
afrawley@susmangodfrey.com  
espanos@susmangodfrey.com

Scarlett Collings (4985602)  
SUSMAN GODFREY L.L.P.  
1000 Louisiana, Suite 5100  
Houston, TX 77002  
Telephone: (713) 651-9366  
Facsimile: (713) 654-6666  
scollings@susmangodfrey.com

Steven Lieberman (SL8687)  
Jennifer B. Maisel (5096995)  
Kristen J. Logan (*pro hac vice*)  
ROTHWELL, FIGG, ERNST & MANBECK, P.C.  
901 New York Avenue, N.W., Suite 900 East  
Washington, DC 20001  
Telephone: (202) 783-6040  
Facsimile: (202) 783 6031



slieberman@rothwellfigg.com  
jmaisel@rothwellfigg.com  
klogan@rothwellfigg.com

*Attorneys for Plaintiff*  
*The New York Times Company*